

2 CHAPTER **Data, Reality, and Problem Solving**

- 2.1 The Lords of Data
- 2.2 Data Classification
- 2.3 Time Series Data vs.
Cross-Sectional Data
- CR Chapter Review

When you encounter data, ask yourself: is the data credible? Consider the following questions:

1. Is the concept under study adequately reflected by the proposed measurements?
2. Is the data measured accurately?
3. Is there a sufficient quantity of the data to draw a reasonable conclusion?

The Scientific Method

1. Gather information about the phenomenon being studied;
2. On the basis of the data, formulate a preliminary generalization or hypothesis;
3. Collect further data to test the hypothesis;
4. If the data and other subsequent experiments support the hypothesis, it becomes a law.

PROCEDURE

Decision-Making Method

1. Clearly define the problem and any influential variables.
2. Decide upon objectives and decision criteria for choosing a solution.
3. Create alternative solutions.
4. Compare alternatives using the criteria established in the second step.
5. Implement the chosen alternative.
6. Check the results to make sure the desired results are achieved.

PROCEDURE

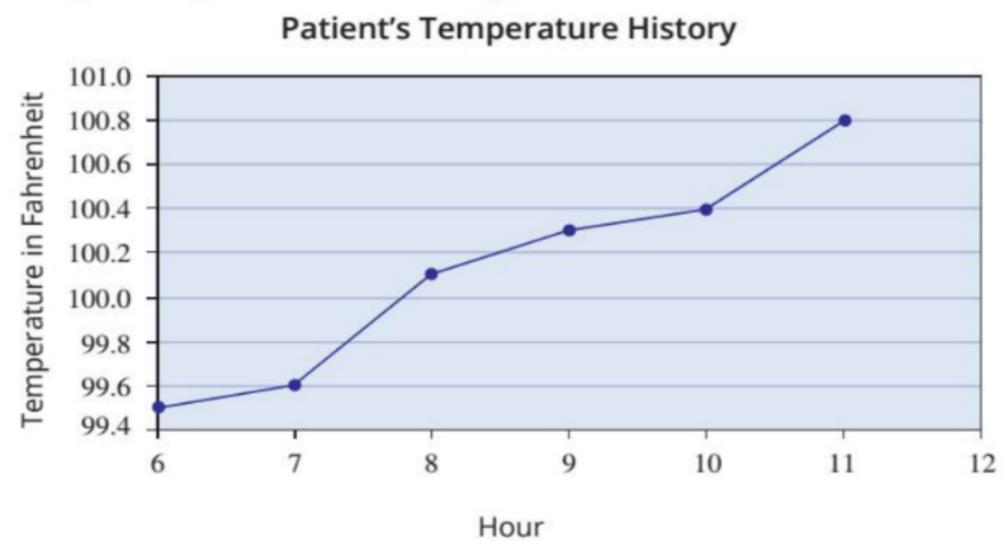
Confounding Variables

Confounding variables are “extra” variables that are not accounted for during experimentation and can cause results to become skewed.

DEFINITION



Figure 2.1.1



Collecting Data

Essentially, there are two ways to obtain data: **observation** and **controlled experiments**.

Response Variable

A **response variable** measures the outcome of interest in a study.

DEFINITION

Explanatory Variable

An **explanatory variable** causes or explains changes in a response variable.

DEFINITION

Suppose a new species of tomato has been genetically engineered to increase yields.
 The question: Does the new species produce higher yields?

Plot 1	Plot 2	Plot 3	Plot 4
Plot 5	Plot 6	Plot 7	Plot 8
Plot 9	Plot 10	Plot 11	Plot 12
Plot 13	Plot 14	Plot 15	Plot 16

- Example 2.1.2

Does an SAT preparation course improve performance on the SAT?

Placebos

A **placebo** is a fake treatment that has the potential to cause a response.

DEFINITION

London scientists conducted a study to determine if chocolate can trigger migraines. Twelve migraine-prone subjects were given a peppermint-laced chocolate candy and eight migraine-prone subjects were given a peppermint-laced placebo made of carob, peppermint, and vegetable fat. Five subjects from the group given chocolate developed a migraine headache within one day. No one from the group given the placebo developed a migraine in the same time period.²

- a. Which phase of the Scientific Method best describes this study?
- b. Is this an observational study or a controlled experiment?
- c. What is the response variable?
- d. What is the explanatory variable?
- e. Which group is the treatment group?
- f. Which group is the control group?

Qualitative and Quantitative Data

Data is sometimes categorized as qualitative or quantitative.

Qualitative

Qualitative data is measured on a nominal or ordinal scale.

DEFINITION

Quantitative

Quantitative data is measured on an interval or ratio scale.

DEFINITION

The rankings of three books on the New York Times Best Sellers List.

Difficulty level: High

Are these data qualitative or quantitative?

list of country codes that you must dial before a telephone number when calling another country. For example, the country code for the United States is 001 and the country code for India is 091.

The heights of 48 randomly selected female students.

Discrete

Data in which the observations are restricted to a set of values (such as 1, 2, 3, 4) that possess gaps is called **discrete**.

DEFINITION

Continuous

Data that can take on any value within some interval is called **continuous**.

DEFINITION

Level of Measurement

The quality of data is referred to as its **level of measurement**.

DEFINITION

The terms used to describe the quality of data are nominal, ordinal,
interval, and ratio.

Quantitative

Qualitative

Nominal

Data that represents whether a variable possesses some characteristic is called **nominal**.

DEFINITION

Ordinal Data

Ordinal data represents categories that have some associated order.

DEFINITION

Frosty Pops taste _____.

1–very bad 2–bad 3–fair 4–good 5–very good

Interval

If the data can be ordered and the arithmetic difference is meaningful, the data is **interval**.

DEFINITION

$$48 \text{ degrees} - 45 \text{ degrees} = 3 \text{ degrees}$$

$$72 \text{ degrees} - 69 \text{ degrees} = 3 \text{ degrees}$$

Ratio Data

Ratio data is similar to interval data, except that it has a meaningful zero value.

DEFINITION

Is money a ratio variable?

Say a friend had \$40 and you had \$20.

$$\frac{\$40}{\$20} = 2$$

According to the ratio we just computed your friend has twice as much money as you. Is this really true? Money is a ratio variable because ratios (quotients) are meaningful. If someone does have \$40 and you have \$20, they do have twice as much money as you.

The number of days traveled last month by 100 randomly selected employees.

Quantitative

Discrete

Ratio

The temperature in Fahrenheit of a U.S. city each day for a month.

Quantitative
Continuous
Interval

A company's international marketing group asked the following question to residents of 50 countries: "What has been your experience with American products?"

1) Below Average 2) Average 3) Above Average 4) Good to Excellent

Qualitative
Neither
Ordinal

A survey response to "what is your favorite color?"

1) Red

2) Blue

3) Green

4) Orange

Qualitative → Neither
Nominal

The results of a study investigating the nutritional status of mid-nineteenth century Americans were reported in "The Height and Weight of West Point Cadets: Dietary Changes in Antebellum America," in the *Journal of Economic History*. The data is based upon physical examination lists for West Point applicants from 1843 to 1894. Some of the information obtained from each cadet were his height, weight, the state from which the cadet was appointed, the occupation of the father, the income of the parents, and the type of home residence (city, town, or rural) of the cadet.

height \rightarrow Quant, C, R
 weight \rightarrow Quant, C, R
 State \rightarrow Quality, N

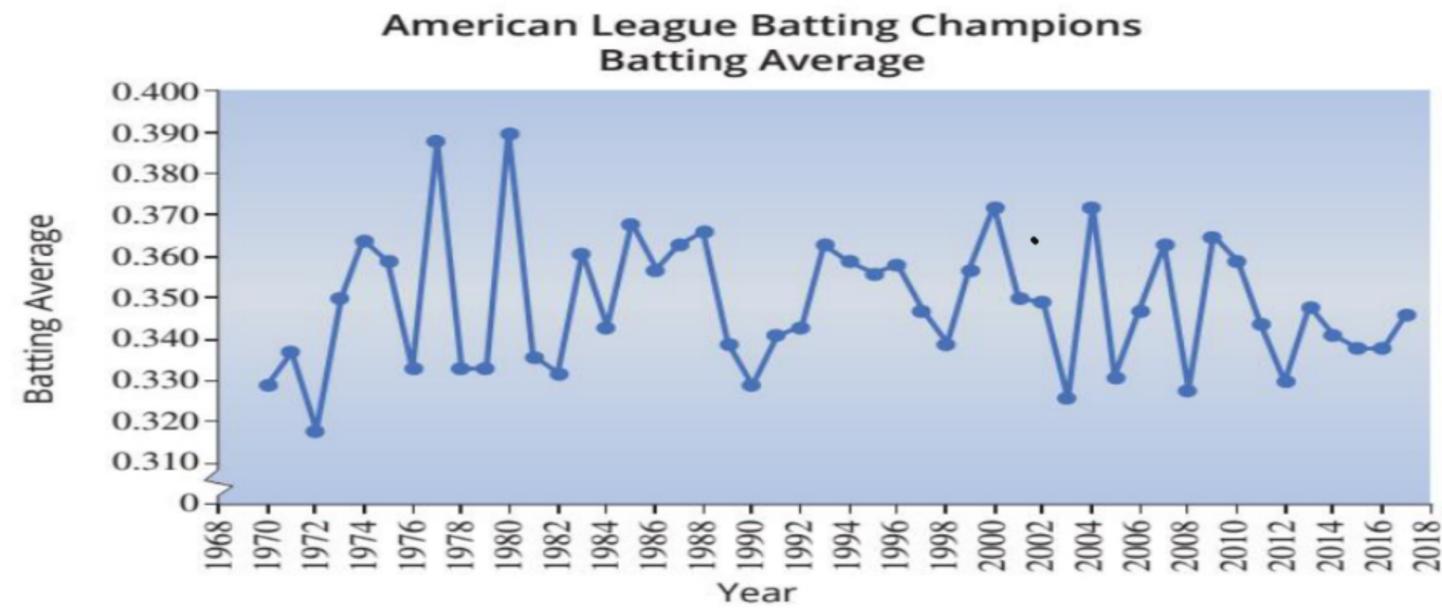
} Occup of father \rightarrow Quality, N
 } Income \rightarrow Quantity, C, R
 } residence \rightarrow Quality, N



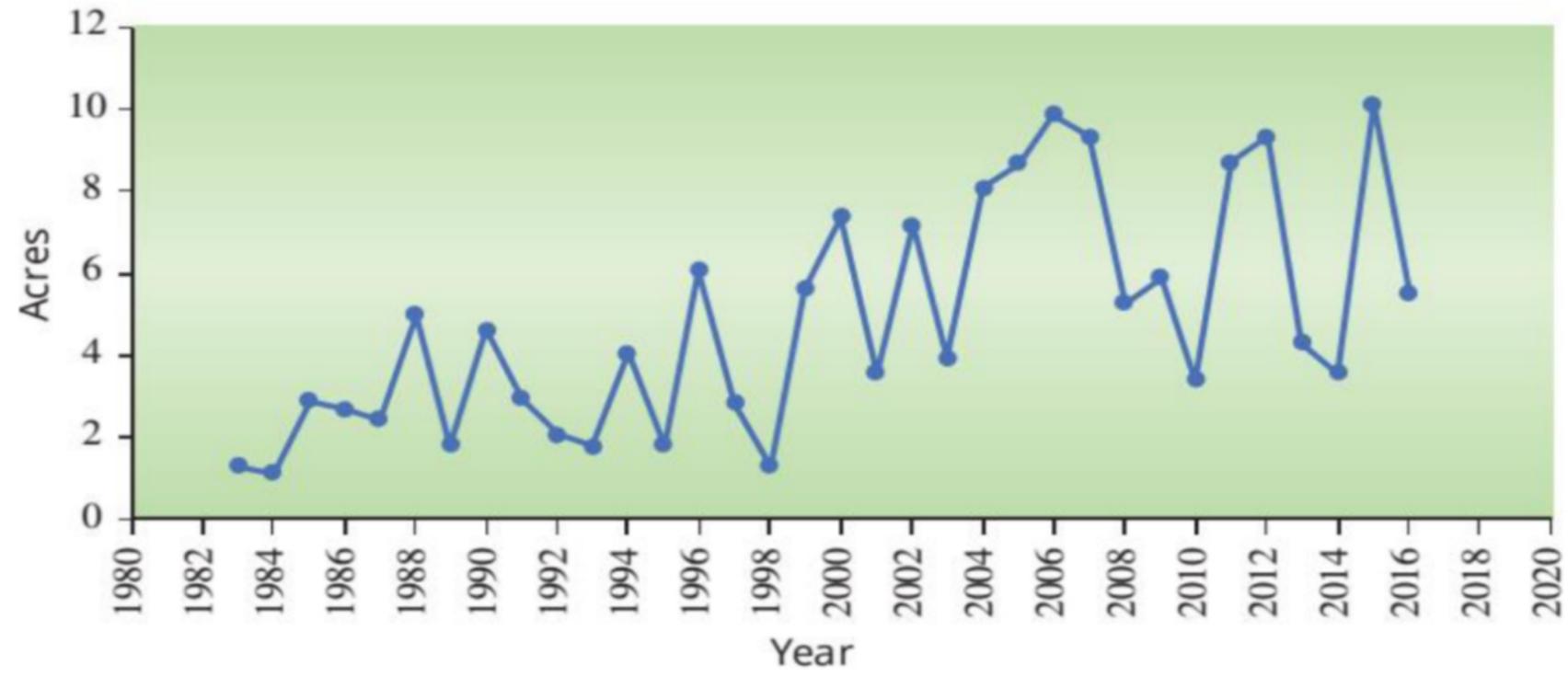
Time Series

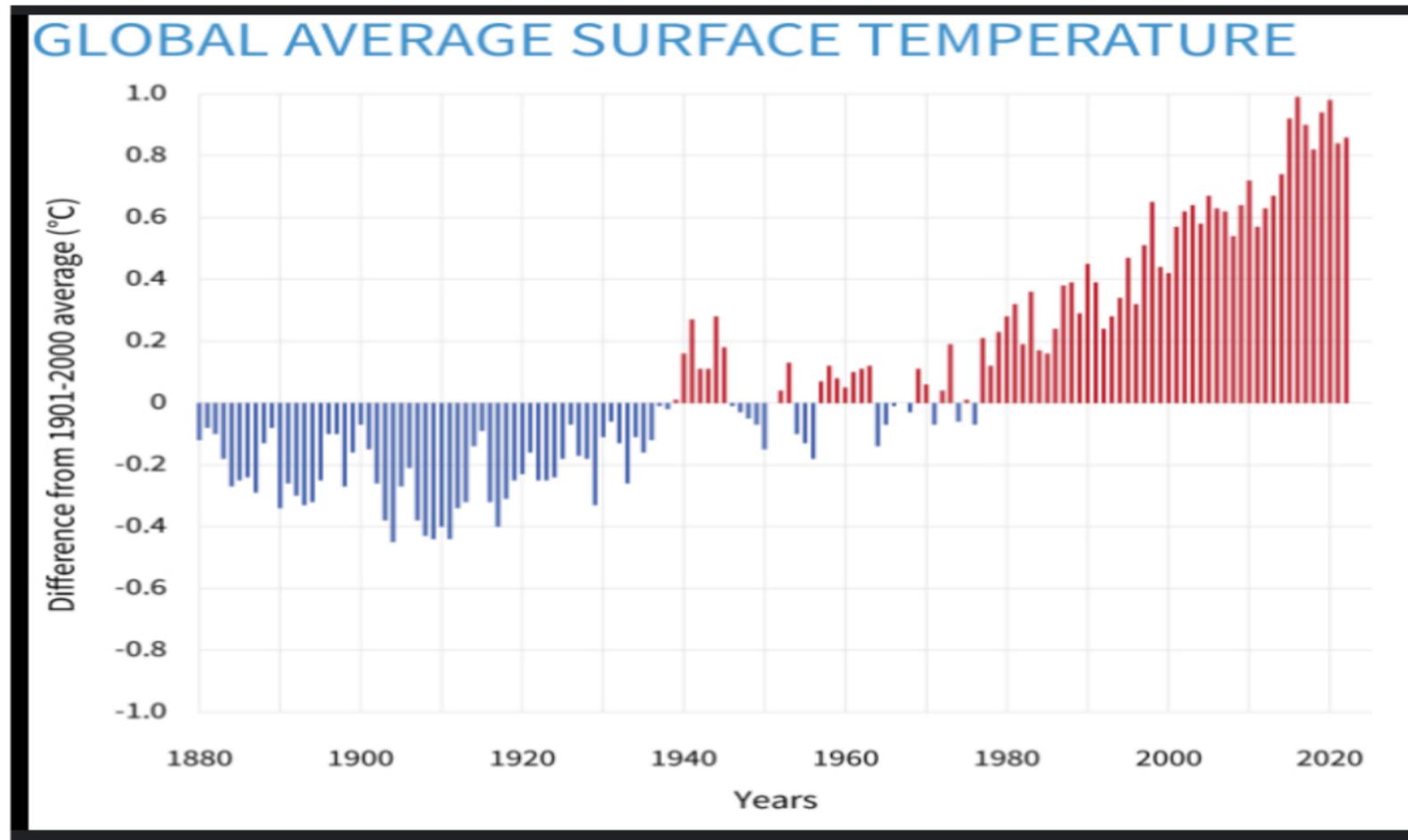
Time series data originates as measurements usually taken from some process over equally spaced intervals of time.

DEFINITION



Acres Burned (in Millions of Acres)





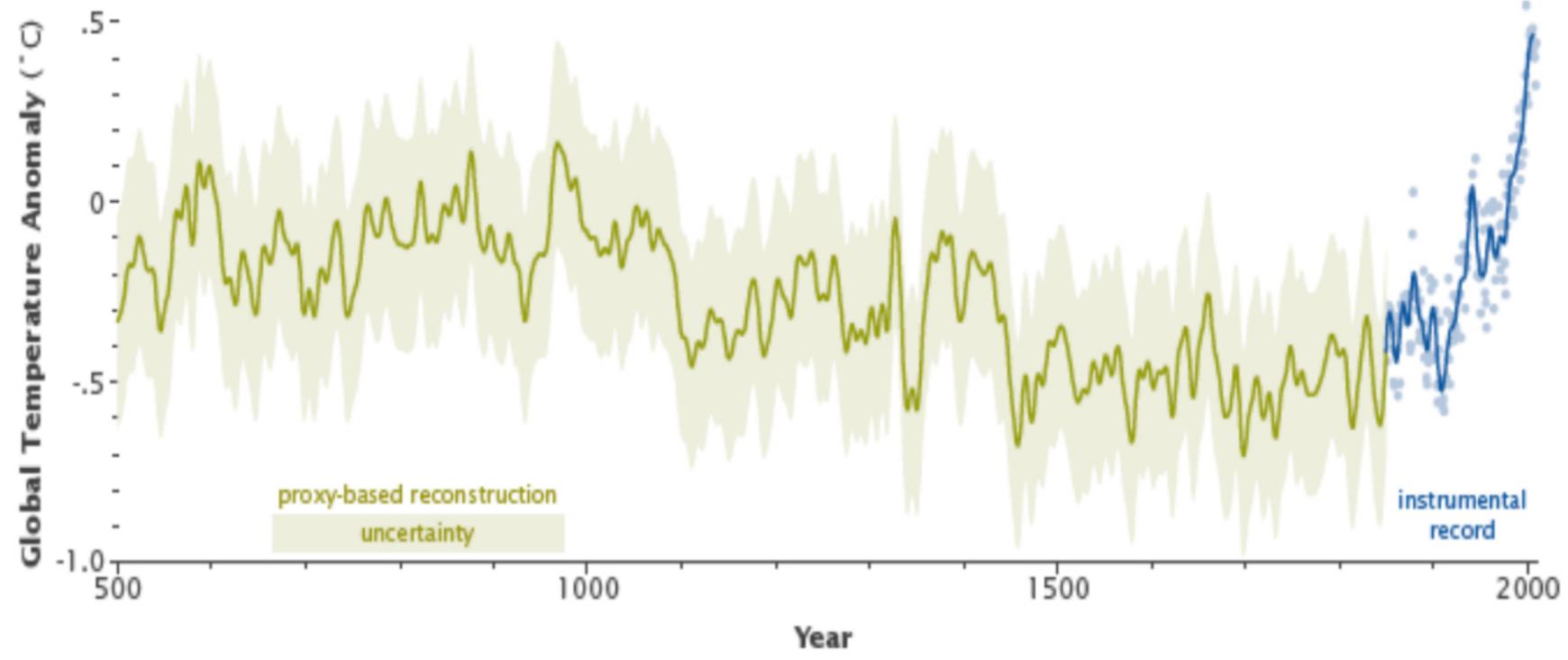


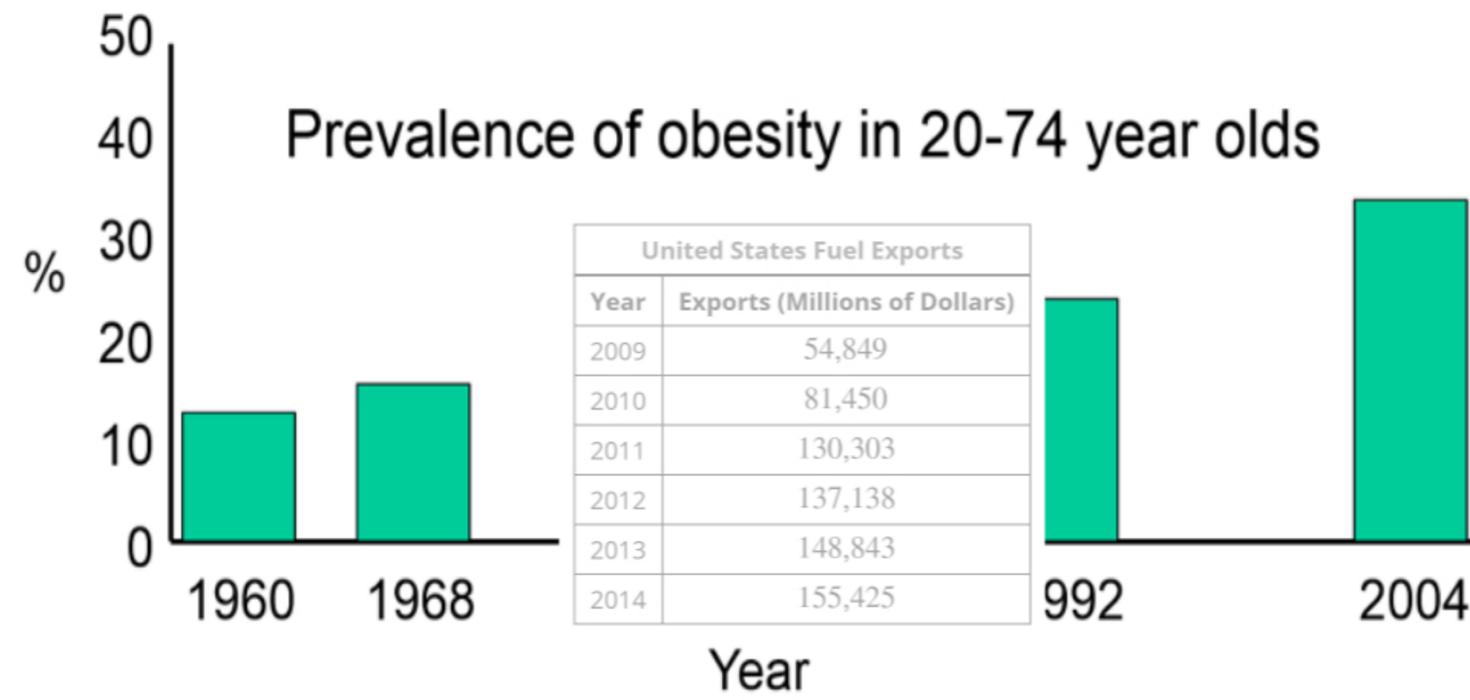
Table 2.3.2 - Life Expectancy at Birth 2015	
Country	Life Expectancy
Afghanistan	61
Australia	83
Botswana	66
Egypt	71
Guatemala	72
Japan	84
Kenya	63
Sierra Leone	46
Spain	83
Sri Lanka	75
Sweden	82
United Kingdom	81
United States	79

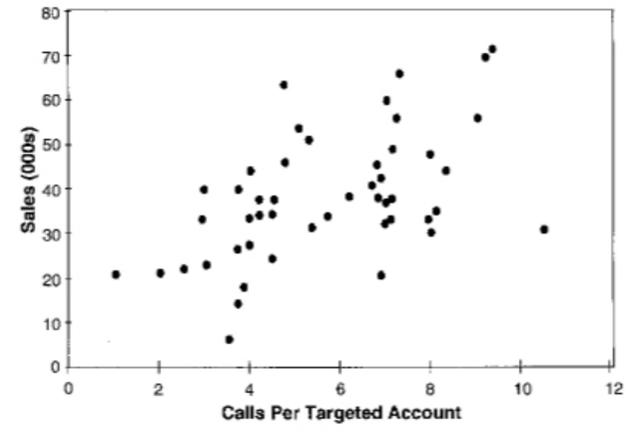
Cross-Sectional Data

Cross-Sectional Data

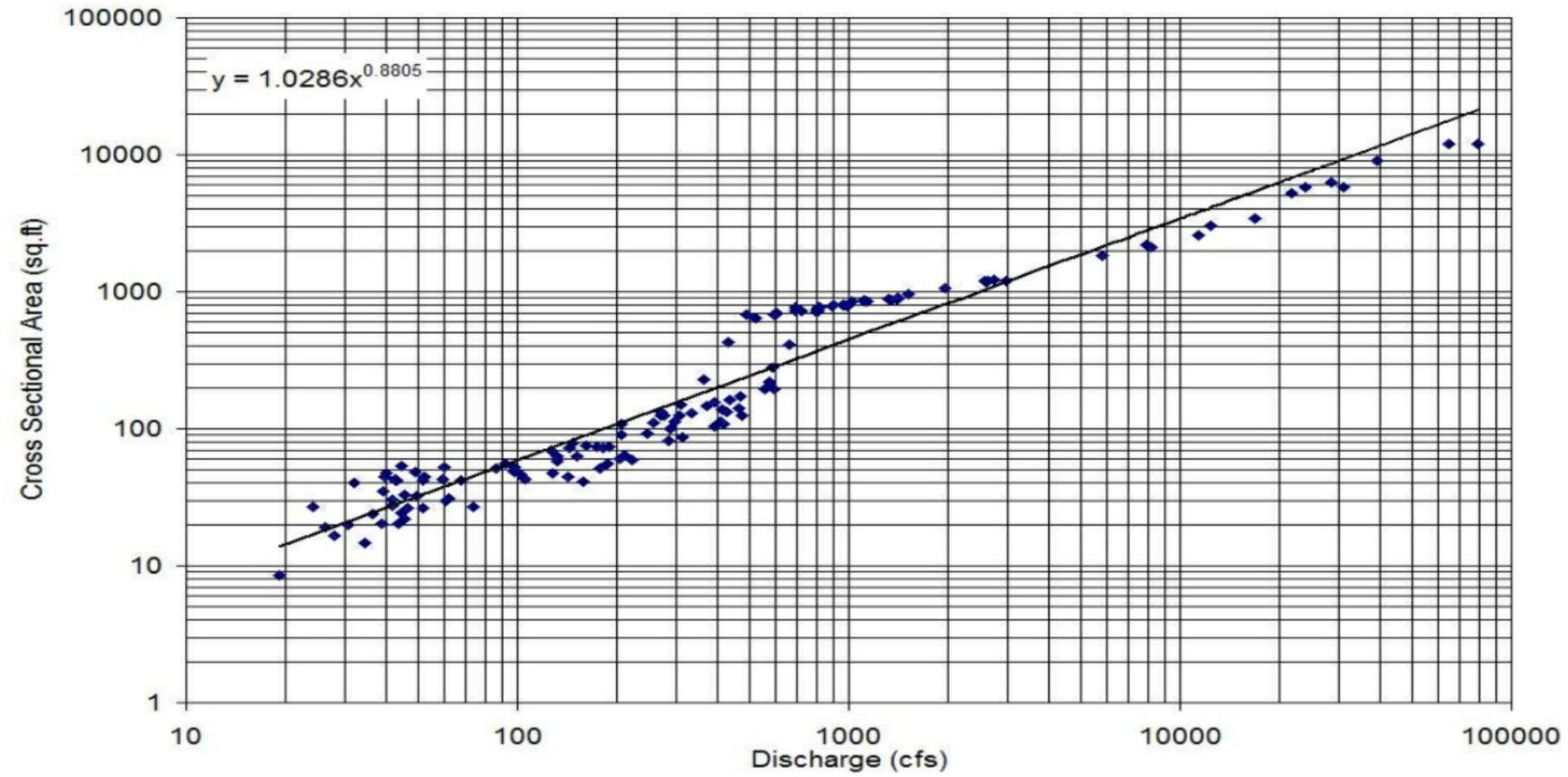
Cross-sectional data are measurements created at approximately the same period of time.

DEFINITION





Cross Sectional Area (sq. ft) vs. Discharge (cfs)



United States Fuel Exports	
Year	Exports (Millions of Dollars)
2009	54,849
2010	81,450
2011	130,303
2012	137,138
2013	148,843
2014	155,425

U.S. Unemployment Rate in 2015	
Month	Unemployment Rate (%)
January	5.7
February	5.5
March	5.5
April	5.4
May	5.5
June	5.3

U.S. Gas Prices by Area: September 2015	
Area	Price (\$)
East Coast	2.337
Midwest	2.354
Gulf Coast	2.12
Rocky Mountain	2.688
West Coast	2.975
West Coast less California	2.697

Sales in 2015: Grocery Stores	
Month	Sales (Millions of Dollars)
January	50,652
February	46,192
March	50,329
April	49,257
May	52,290
June	50,073

2013 U.S. Household Income and Benefits	
Income Range	Number of Households
Less than \$10,000	8,380,364
\$10,000 to \$14,999	6,214,548
\$15,000 to \$24,999	12,468,604
\$25,000 to \$34,999	11,929,761
\$35,000 to \$49,999	15,723,148
\$50,000 to \$74,999	20,744,045

